

University of Groningen

Ordered latent class models in nonparametric item response theory

van Onna, Maria Johanna Henrica

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2004

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

van Onna, M. J. H. (2004). *Ordered latent class models in nonparametric item response theory*. s.n.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Summary

A ordered latent class model can be interpreted as a nonparametric item response theory model. This offers the possibility to estimate and check a nonparametric item response theory model in new ways. In addition, the extent to which ordered latent class models can be used to estimate the sampling distribution of scaling coefficient H is investigated in this thesis.

Item response theory (IRT) deals with the statistical modelling of responses of persons to tests. A nonparametric item response theory (NIRT) model has the assumptions of unidimensionality (UD), local independence (LI), and monotonicity (M). The assumption of UD says that only one single latent trait influences the response probabilities. Local independence means that the response probabilities of different items are independent of each other given a position on the latent trait. The assumption of monotonicity deals with the shape of the item response function (IRF). This IRF is the probability of a positive response as a function of the latent trait. Monotonicity supposes that the IRF is non-decreasing in the latent trait. These three assumptions together ensure that the total score on a test can be used to rank order subjects in accordance with their positions on the latent trait. A fourth common assumption in NIRT is the assumption of non-intersection (NI). It states that IRFs of different items do not intersect. The advantage is that the ordering of the item difficulties is the same for each position on the latent trait.

The assumptions of an ordered latent class model are comparable to those of a NIRT model. The latent classes are ordered on one single dimension (UD). The conditional response probabilities are independent given a latent class (LI). Order restrictions on the conditional response probabilities may ensure assumptions like M and NI. A difference with the description of NIRT

above is that the latent trait is not continuous, but divided in a restricted number of latent classes.

The items may not be dichotomous, but polytomous. The number of order restrictions increases fast with the number of response categories per item. In Chapter 2, a Bayesian estimation algorithm for an ordered latent class model for polytomous items is presented. In addition, several Bayesian model selection methods are applied. The Gibbs sampler, a Markov chain Monte Carlo method, can sample from the multivariate posterior distribution of the parameters under the imposed restrictions. Bayesian model selection can be done with posterior predictive checks and with Bayes factors. Several discrepancy functions can be designed for a posterior predictive check. Each of them may be sensitive to a different aspect of the data, like frequencies of entire response vectors, or response frequencies on two items.

Two aspects are involved in model selection: the strictness of the assumptions, and the number of latent classes. The first aspect asks whether M and NI can be imposed on the data. In Chapter 2, a simulation study is described in which the application of Bayesian methods to ordered latent class models is evaluated. Also, existing data with polytomous items are analyzed, as an example. The conclusion is that the Bayesian estimation procedure functions well under the order restrictions and the large amount of parameters. The Bayesian model selection methods are sensitive to the assumption level, but much less so to the number of latent classes.

If a model is rejected for the data, a detailed search for the cause of this rejection is advisable. If, for example, only one item violates the assumption of M , it is more profitable to remove this item from the scale than to reject the entire scale. In Chapter 3, two NIRT methods (a descriptive and an inferential method) for the detection of general and specific model misfit are studied using a simulation study. The descriptive method is based on the work of Mokken (1971). Descriptive statistics at scale level or at item level can indicate misfit if they surpass some boundary values. The inferential method is based on posterior predictive checks of the ordered latent class model. The discrepancy functions are defined both at scale level as at item level.

In Chapter 3, it is found that strict model testing at scale level is best

done by an ordered latent class approach. However, an inspection of all items individually gives a much clearer image of possible model misfit. Within the descriptive method, which is much faster than the inferential method, the $crit_i$ -value is best fit to detect decreasing IRFs and discern them from flat IRFs, which are a borderline case of M. Flat IRFs are not useful for rank ordering persons, and are detected as well by the descriptive index H_i .

Methods within NIRT for the detection of multidimensionality are studied in Chapter 4. Several kinds of multidimensionality are simulated in the study. In a factor analytic interpretation of multidimensionality, the dimensions function compensatory. This means that a low position on one dimension can be compensated for by a high position on another dimension, to obtain a high response probability after all. A specific case of a compensatory model is a simple structure model, in which each item loads only on one dimension. In this case, the items can be clustered in unidimensional sets of items. Multiple groups of respondents may exist, in which the relation between latent trait to be measured and the response probabilities (the IRF) differs. This case is called unobserved heterogeneity. The assumption of UD is violated here as well.

The four methods which are studied in Chapter 4 are the DETECT-algorithm and accompanying DETECT-index, the cluster algorithm of MSP, information criteria which compare a unidimensional latent class factor model with a two-dimensional latent class factor model, and posterior predictive checks of the (unidimensional) ordered latent class model. The four methods showed large correspondences in detecting violations of UD. The multidimensionality in samples which had simple structure was detected. However, two completely compensatory latent traits and unobserved heterogeneity on one out of ten items were not detected.

The non-detection was not disturbing in the case of unobserved heterogeneity, because the ordering of respondents on the total score turned out to be robust against this violation of UD. However, in the case of compensatory latent traits, the ordering of respondents on the total score did not match the ordering on the first dimension, which was the trait intended to be measured. The ordering on the total score did match the ordering on the total

test dimension, which is the sum of the first and second dimension. The often used reliability index Cronbach's alpha is an indicator of the association between the total score and the total test dimension, but it does not reveal the dimensional structure of the data.

In NIRT, coefficient H is a scaling index of a set of items. It indicates to which extent respondents can be ordered by means of the items. The sampling distribution of coefficient H gives information about the variation of H in repeated samples. However, the sampling distribution of H is only derived asymptotically and under restrictive conditions. An alternative to estimate the variance of H are bootstrap methods. An impression of the stability of H can also be obtained by estimating an ordered latent class model. In Chapter 5, the different methods are compared to each other in a simulation study.

The asymptotic variance of H was very close to the sampling distribution, despite samples of only 200 respondents and the restriction of a fixed item order. The naive bootstrap also resulted in a good estimate of the sampling distribution. The advantage of the methods based on an estimated latent class model was that it was indicated when the data did not fulfill the NIRT assumptions. Of the methods based on an estimated ordered latent class model, the plug-in interval on the basis of the ML-estimate (parametric bootstrap) reflected the sampling distribution best, when the NIRT assumptions were fulfilled. The location of the interval, however, was a little too high. A remarkable conclusion is that a model with only five latent classes was able to approximate a continuous latent trait sufficiently. More latent classes did not change the results.